AN ELEMENTARY EXPOSITION OF THE THEOREM OF BERNOULLI WITH APPLICATIONS TO STATISTICS

By Professor H. L. RIETZ, University of Iowa

In connection with the wide use of certain statistical methods within the past few years, there is coming to be some recognition of the importance of establishing a measure of the degrees of confidence that can properly be placed in inferences from statistical results such as mean values, standard deviations and coefficients of correlation. This recognition is shown in the increased application of probable errors, even if the applications are in many cases made without knowledge of the derivations or limitations of the formulas employed. The need for such criteria in passing judgment on the significance of the simplest of statistical results may perhaps be made clear by an appropriate illustration.

In the third edition of American Men of Science by Cattell and Brimhall, it is stated that a group of "scientific men reported 716 sons and 668 daughters." The valid inference is drawn that the "difference falls within the limits of chance variation, and is not likely to be significant." On the same page, 804, we find that a group of scientific men report 1705 brothers and 1527 sisters. These data suggest the following questions of simple statistical sampling: What is the probability in throwing 1384 coins that the number of heads will differ from $\frac{1384}{2} = 692$ by as much or more than $692 - 668 = 24$? What is the probability in throwing 3232 coins that the number of heads will differ from $\frac{3232}{2} = 1616$ by as much or more than $1705 - 1616 = 89$? The answers to these questions are obtained by an application of what is known in certain important mathematical literature* as the theorem of Bernoulli, although the theorem in the form in which we shall use it contains much in addition ** to the Bernoulli theorem as it appeared in the latter's works.

* See German and French Encyclopedias of Mathematics—Papers on Probability.

** See Laplace, Theorie Analytique des Probabilities, Introduction, p. XLVII, Bertrand, Calcul des Probabilities, Chapter V.

THEOREM OF BERNOULLI

We shall assume that during a set of s trials, the probability of the happening of an event is a constant p from trial to trial, and that

$$p + q = 1.$$

Then the probabilities that the event will happen exactly s, $s - 1$, $s - 2$, $\cdot \cdot \cdot$, 1, 0 times in s trials are given by the successive terms of the *binomial* expansion

$$(p + q)^s = p^s + sp^{s-1}q + \qquad + \frac{s!}{m!(s-m)!} p^m q^{s-m} + \cdots$$
$$+ spq^{s-1} + q^s, \qquad (1)$$

To find the "most probable" number of happenings, we seek the value of m to which a maximum term of (1) corresponds. It is easly shown that

$$m = ps$$

gives the maximum term if ps is an integer. When ps is not an integer, the integer m is such that

$$ps - q \gtreqless m \gtreqless ps + p$$

gives a most probable value.

When $ps - q$ and $ps + p$ are integers, there occur two equal terms in (1) each of which is larger than any other term of (1). For example, note the equality of the first and second terms of the expansion of

$$(\tfrac{5}{6} + \tfrac{1}{6})^5.$$

For the present, let us assume that ps is an integer, and let us represent the terms of (1) by ordinates of the curve

$$y_x = f(x),$$

where x marks deviations from the maximum term as an origin. Then we have

$$y_x = \frac{s!}{(ps - x)!(qs + x)!} p^{ps+x} q^{qs-x}, \quad . \quad . \quad . \quad (2)$$

and

$$y_{-x} = \frac{s!}{(ps - x)!\,(qs + x)!} p^{ps-x} q^{qs+x}, \qquad (3)$$

By making $x = 0$ in (2) or in (3), we have the maximum ordinate

$$y_o = \frac{s!}{ps!\,qs!}\, p^{ps}q^{qs}.$$

With values of s that are reasonably large for statistical purposes, it is usually impractical to calculate the factorial in (2) and (3) without some special methods of approximation. Such a method is provided by an application of Stirling's theorem for the representation of large factorials.

This theorem states that*

$$s! = s^{s+\frac{1}{2}}e^{-s}\sqrt{2\pi} \text{ approximately.} \tag{4}$$

The substitution of this value for $s!$ and corresponding values for $ps!$ and $qs!$ in

$$y_o = \frac{s!}{ps!\,qs!}\, p^{ps}q^{qs}$$

gives, after some simplification,

$$y_o = \frac{1}{\sqrt{2\pi spq}},$$

To illustrate, the most probable value in throwing 1000 coins, namely 500 heads and 500 tails, has a probability

$$y_o = \frac{1}{\sqrt{500\pi}} = .02523.$$

It is important to note that this most probable value is not likely to be obtained in a single trial since its probability is only a little more than $\frac{1}{40}$. It may be of interest to the reader to compare the simplicity of the calcuation of y_o as above with the calculation of

$$\frac{1000!}{500!\,500!}\left(\frac{1}{2}\right)^{1000} \text{ by logarithms.}$$

By the application of Stirling's theorem to (2), we obtain, after slight simplification,

$$y_x = \frac{1}{\sqrt{2\pi spq}}\left(1+\frac{x}{ps}\right)^{-ps-x-\frac{1}{2}}\left(1-\frac{x}{qs}\right)^{-qs+x-\frac{1}{2}} \tag{5}$$

* For proof, see Whittaker and Watson, Modern Analysis, Third Edition, p. 253; Czuber, Wahrschenlichkeitsrechnung, I, 1908, p. 22.

In practical statistical problems, we are in general interested in deviations, x, that are small compared to the most probable value ps. In fact, we may well confine our attention to deviations that are of the order of $\sqrt{s}$ when we are discussing fluctuations in sampling. With this limitation on x, it is shown in the appendix to this paper that the sum

$$y_x + y_{-x} = \frac{2}{\sqrt{2\pi pqs}}\, e^{-\frac{x^2}{2pqs}} \text{ approximately.} \tag{6}$$

Since we very commonly make our inquiries about a given deviation on either side of the most probable, we are interested in the sum $y_x + y_{-x}$ given by (6).

The limitation that ps be an integer may well be removed. From what was shown above about the most probable value, we may in any case write for the most probable number of happenings

$$m = ps + k$$

where $-1 < k < +1$.

With larger values of s, it can be shown that the difference brought about by the use of $ps + k$ instead of ps is of negligible importance in our problem of fluctuations in sampling.

We are particularly interested in finding the probability that the variable deviation x will remain within assigned bounds, say within d and $-d$ inclusive. To find this probability requires a method of finding the sum

$$y_d + y_{d-1} + \cdot \cdot \cdot + y_1 + y_0 + y_{-1} + \cdot \cdot \cdot + y_{-d} = \sum_{x=-d}^{x=d} y_x. \tag{7}$$

As there is likely to be a large number of y's, some special method of finding the sum is of practical value. The sum of such a large set of ordinates may be found by the Euler-Maclaurin formula of the calculus of finite differences. For the purpose of a sampling problem, a simpler method of approximation than that provided by the Euler-Maclaurin Theorem is suitable. It will be convenient in what follows to make

$$y'_x = \frac{y_x + y_{-x}}{2} = \frac{1}{\sqrt{2\pi pqs}}\, e^{-\frac{x^2}{2pqs}} \qquad (8)$$

We may now well conceive of obtaining the approximate sum of the ordinates in (8) by finding the area enclosed by the curve, the x-axis, and the ordinates $x = -d - \frac{1}{2}$, and $x = d + \frac{1}{2}$.

For this purpose, we use as an approximate value of the area, the integral,

$$\int_{-d-\frac{1}{2}}^{d+\frac{1}{2}} y'_x dx = 2 \int_{0}^{d+\frac{1}{2}} y'_x dx$$

$$= \frac{2}{\sqrt{2\pi pqs}} \int_{0}^{d+\frac{1}{2}} e^{-\frac{x^2}{2pqs}}\, dx \qquad (9)$$

The theorem of Bernoulli may now be stated by saying, (1) *that ps is a most probable value,* (2) *that formula* (9) *gives the probability that a deviation x from the most probable will not exceed an assigned deviation d on either side of the most probable.*

The numerical value of (9) is readily obtained in any particular case by the use of the normal probability integral as we shall show by applications to the numerical questions proposed on pp. 1 and 2.

In the first question,

$$s = 1384,$$

$$pqs = 346,$$

$$\sqrt{pqs} = 18.601,$$

$$d = 716 - 692 = 24,$$

$$\frac{d + \frac{1}{2}}{\sqrt{pqs}} = 1.3171.$$

From a table of the normal probability integral (Table IV, Davenport, Statistical Methods), we find for the deviation $\frac{x}{\sigma} = \frac{d + \frac{1}{2}}{\sqrt{pqs}} = 1.3171$, the value of (9) to be $P = .8122$.

This is the probability in throwing 1384 coins that the deviation of the number of heads from $\frac{1384}{2} = 692$ will not exceed 24. The probability of a deviation greater than 24 is then $1 - .8122 = .1878$. Expressed in another way, we may say we should predict that, in the long run, a deviation greater than 24 on either side of the most probable will occur slightly less than once per five trials. In the second question,

$$s = 3232,$$
$$\sqrt{pqs} = 28.425,$$
$$d = 1705 - 1616 = 89,$$
$$\frac{d + \frac{1}{2}}{\sqrt{pqs}} = 3.1486.$$

Referring now to a table of the normal probability integral, we find for the deviation

$$\frac{x}{\sigma} = \frac{d + \frac{1}{2}}{\sqrt{spq}} = 3.1486,$$

the value of (9) to be

$$P = .99836.$$

The probability of a deviation greater than the assigned deviation on either side of the most probable $\frac{3232}{2} = 1616$, is then $1 - .99836 = .00164$.

In other words, we predict a deviation larger than 89 on either side of the most probable should occur in the long run about once per $\frac{1}{.00164}$ trials, or roughly once per 600 trials. **We thus have** a quantitative criterion to judge of the significance of the given deviation compared to fluctuations in simple sampling.

By dividing the frequencies under consideration by the total numbers involved, we have relative frequencies, and the theorem of Bernoulli may be stated in terms of these relative frequencies. We should then regard the theorem as furnishing a criterion for testing whether the deviation of a statistical ratio from an assumed probability can be reasonably regarded as a fluctuation in sampling.

To summarize, we may say that the theorem of Bernoulli states (1) the number of happenings that is most probable, (2) the probability, in making s trials with constant probability p, that

the departure of the number of successes from sp will not exceed a given number, or that the departure of a statistical ratio from a known constant probability p will not exceed a given value.

The converse theorem to that of Bernoulli is of great importance in statistics. That is, to determine the probability that an unknown probability of an event will not deviate more than an assigned value from a statistical ratio is a problem of much interest. We shall not attempt here to give the reasoning by which the converse is established because of limitations of space and because the purpose of this paper is accomplished by giving a view of the method of treating one of the simplest problems of fluctuations in sampling.

APPENDIX

Given the function

$$y_x = \frac{1}{\sqrt{2\pi spq}}\left(1+\frac{x}{ps}\right)^{-ps-x-\frac{1}{2}}\left(1-\frac{x}{qs}\right)^{-qs+x-\frac{1}{2}} \qquad (1')$$

marked (5) above, to show that

$$y_x + y_{-x} = \frac{2}{\sqrt{2\pi pqs}}\, e^{-\frac{x^2}{2pqs}} \text{ approximately} \qquad (2')$$

under the limitations on x stated on p. 5.

From (1′)

$$\log\ y_x = -\tfrac{1}{2}\log\ (2\pi pqs) - (ps+x+\tfrac{1}{2})\log\ (1+\tfrac{x}{ps})$$
$$-(qs-x+\tfrac{1}{2})\log\ (1-\tfrac{x}{qs}).$$

By expanding $\log\ (1+\frac{x}{ps})$ and $\log\ (1-\frac{x}{qs})$ in series, and simplifying, we have

$$\log\ y_x = -\tfrac{1}{2}\log.\ (2\pi pqs) + \frac{(p-q)x}{2pqs} - \frac{x^2}{2pqs}$$
$$+\frac{x^3}{6p^2s^2} - \frac{x^3}{6q^2s^2} + \qquad (3')$$

Hence,

$$y_x = \frac{1}{\sqrt{2\pi spq}}\, e^{\frac{(p-q)x}{2pqs} - \frac{x^2}{2pqs} + \frac{x^3}{6p^2s^2} - \frac{x^3}{6q^2s^2} + \cdots}$$

$$=\frac{1}{\sqrt{2\pi spq}}\,e^{-\frac{x^2}{2pqs}}\left(1+\frac{(p-q)x}{2pqs}+\frac{x^3}{6q^2s^2}-\frac{x^3}{6p^2s^2}+\cdots\right) \qquad (4')$$

Similarly,

$$y_{-x}=\frac{1}{\sqrt{2\pi spq}}\,e^{-\frac{x^2}{2pqs}}\left(1-\frac{(p-q)x}{2pqs}-\frac{x^3}{6p^2s^2}+\frac{x^3}{6q^2s^2}\cdots\right) \qquad (5')$$

In the applications to statistics, we are generally interested in deviations, x, that are small compared to ps and qs. In fact, we may well confine our attention in the treatment of fluctuations in sampling to values of x not exceeding $\sqrt{s}$ in order of magnitude. Thus, in (4′) we have not retained in the parenthesis the term $\frac{(p-q)^2x^2}{2(2pqs)^2}$. Under this limitation on x, the sum of terms beyond those written in (4′) and (5′) may be taken as negligible for our purposes.

Hence, we have from (4′) and (5′),

$$y_x+y_{-x}=\frac{2}{\sqrt{2\pi pqs}}\,e^{-\frac{x^2}{2pqs}} \text{ approximately.}$$